

Open Source Philosophy E-Scholarship

Some thoughts on the importance of open source and open access
for emerging digital scholarship

Dr. Stefan Gradmann
Universität Hamburg / Regionales Rechenzentrum
stefan.gradmann@rrz.uni-hamburg.de
www.rrz.uni-hamburg.de/RRZ/S.Gradmann

Presentator's Context: Humanities Computing, DLs & OA

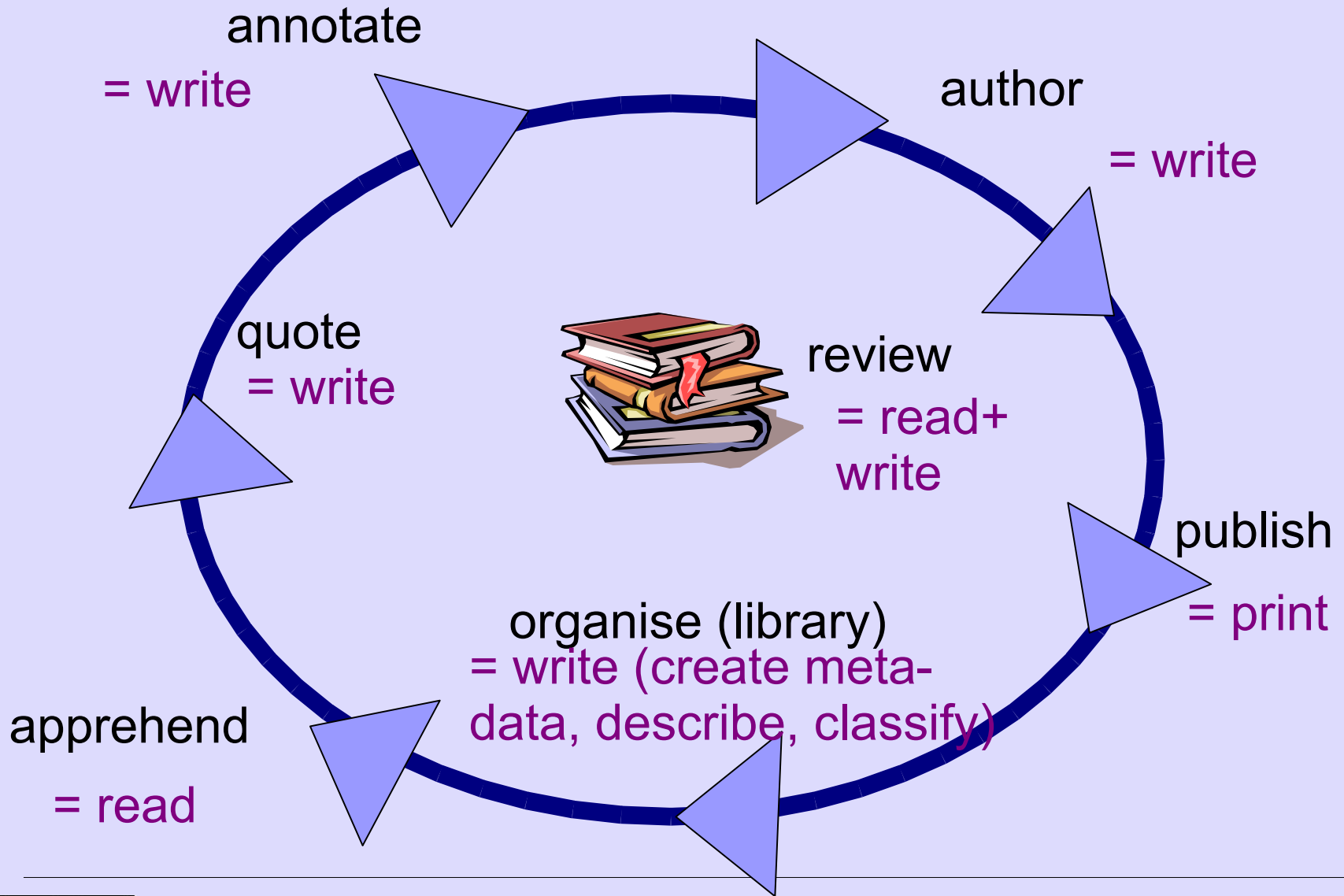
- Not a Wittgenstein scholar, not even a philosopher strictly speaking
- Background in literary scholarship (work on Joyce, Kafka, Arno Schmidt, Greek Mythology) and librarianship
- Hamburg University / Computing Center + HDH
- Open Access (GAP, DINI)
- COST A32 „Scholarly Communities on the Web“, representing Germany in Management Board and leading Technical WG
- International advisor to “Our Cultural Commonwealth” (American Council of Learned Societies Report on Cyberinfrastructure for the Humanities and Social Sciences)
- Closely following RTP Doc (French work on document models)
- Building the European Digital Library / EDLnet: leading WP2 (Interoperability) together with Makx Dekkers (DCMI)

- Two cultures: E-Science vs. E-Scholarship
- Scholarly/scientific information continuum leaving 'emulation mode'
- What specific consequences does this create for digital scholar publication and open document models and standards?
- What is e-scholarship and how does it specifically relate to its object corpora / sources?
- What consequences result from this specific difference regarding 'open source' approaches?
- Some concluding words, including aspects of quality and reputation in open humanities
- **Caveat:** lots of implicit information some of which may not be widespread in the community present in Kirchberg – however, trying to properly define controversial terms such as Open Source or Open Access would take hours!

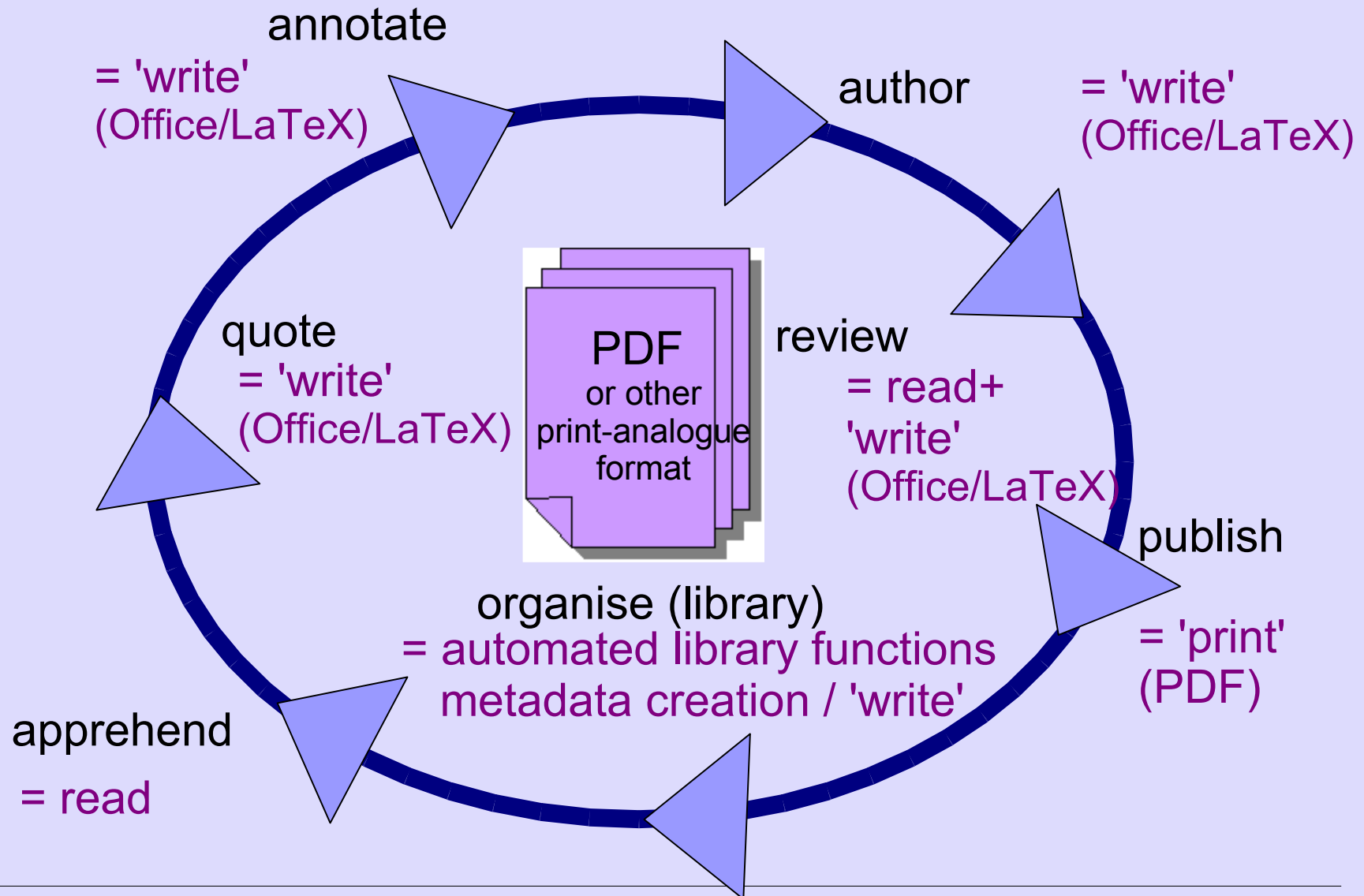
The two scientific cultures: hermeneutical vs. empirical heuristics

- **'Hard' Sciences - Empirical focus**: finding routes towards a known target / intelligent retrieval strategies – “Explain”!
- **Humanities - Hermeneutical focus**: strive for 'knowledge' / finding 'reasons' – “Understand”!
- *„Human interpretation is the heart of the humanities ... devising computer-assisted ways for humans to interpret more effectively vast arrays of the human enterprise is the major challenge.“*
(Michael Jensen in a hearing of the ACLS Commission on Cyberinfrastructure for the Humanities and Social Sciences)
- The division used here is simplistic and inappropriate in many respects – it can be useful, however, to understand some of the specific aspects of Open Access in the context of e-scholarship.
- Starting from W. McCarty: *„Academic publishing is one part of a system of highly interdependent components. Change one component [...] and system-wide effects follow. Hence if we want to be practical we have to consider how to deal with the whole system.“*

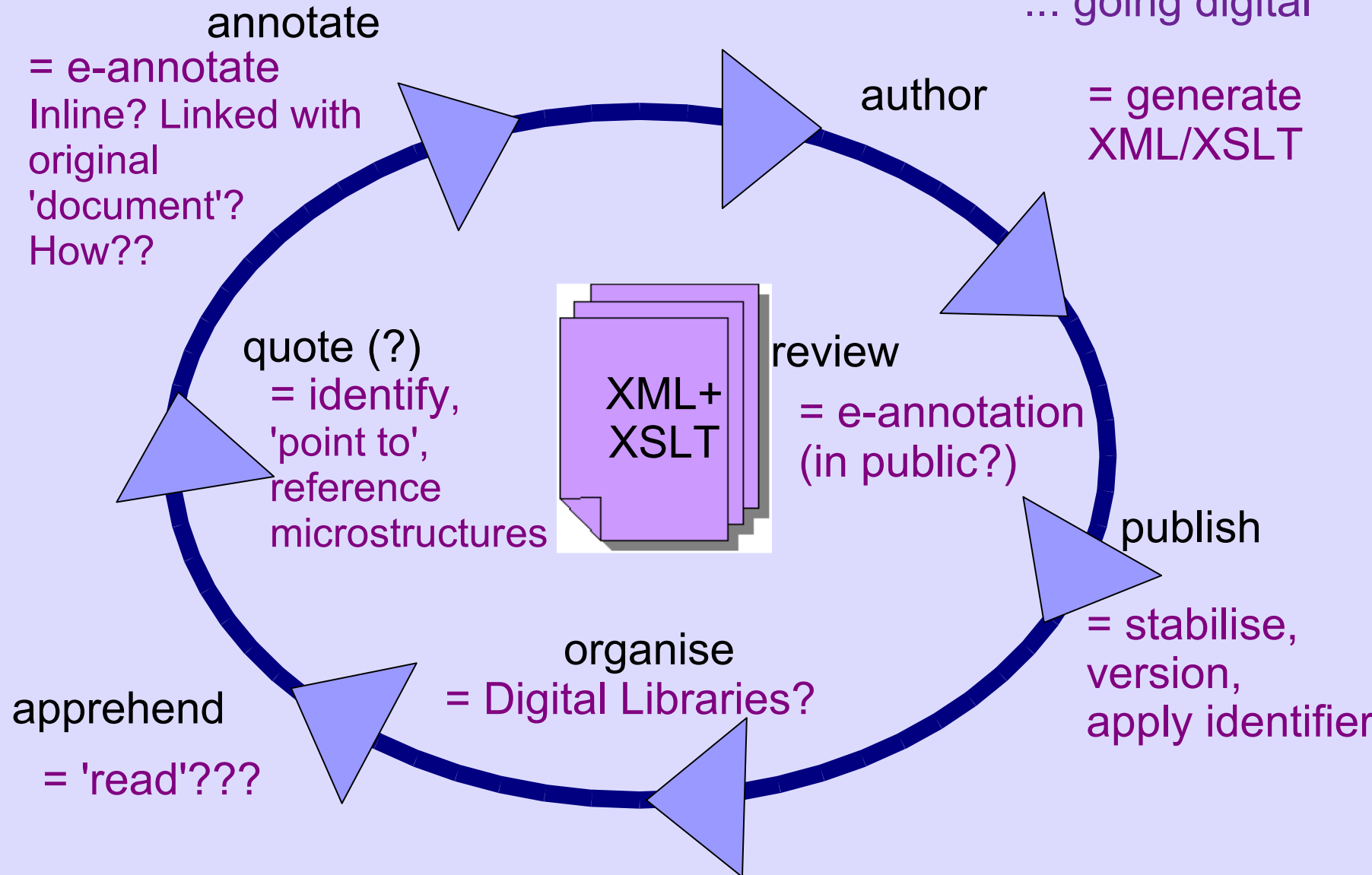
Linear Information Continuum using traditional cultural techniques



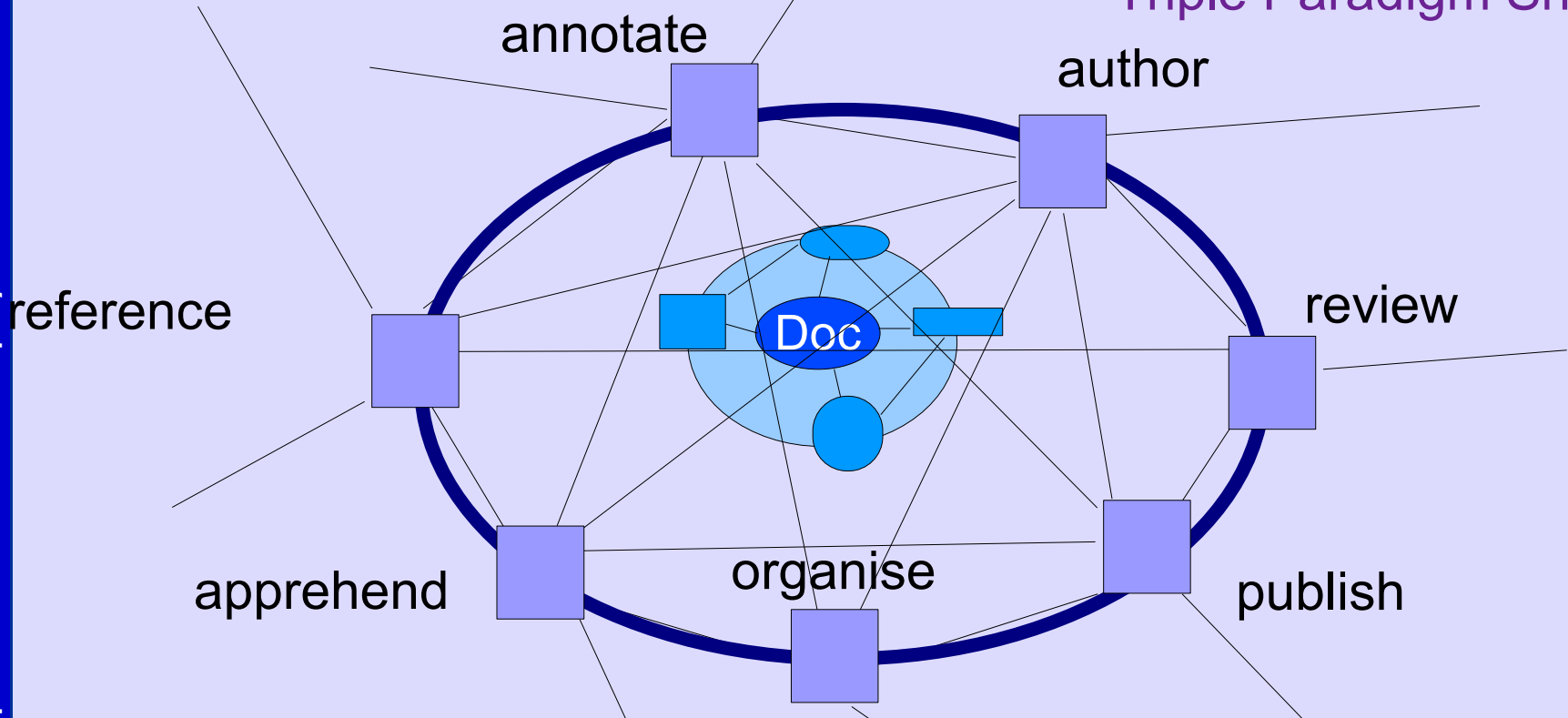
Linear Information Continuum emulating traditional cultural techniques



Linear Information Continuum ... going digital



Scientific Information Continuum Triple Paradigm Shift



- Erosion of the linear / circular function paradigm
- De-Construction of the 'document' notion in a digital, networked context
- Functionality is not any more entirely determined by traditional cultural techniques and related metaphors – and cannot not yet be determined by digital and still emerging technology

Consequences of the triple paradigm shift for the humanities

- The **Erosion of the linear / circular function paradigm** only marginally affects the humanities because of their 'monolithic' publication culture.
 - Journal publications as well as the related workflows and peer reviewing scenarios still play a less prominent role in our context.
 - Most authors in the humanities still basically work in isolated, autonomous settings, group authoring scenarios still tend to be exceptional.
- The **decrease of functional determination by traditional cultural techniques** does affect the humanities in many respects - none of these, however, being specific for the humanities.
- The **De-Construction of the 'document' notion** in digital, networked settings vitally affects the humanities in that it fundamentally changes the conditions of production and use of 'documents' and namely
 - Conditions of **apprehension** and **reuse**
 - Fundamental **signification modes of 'documents' seen as complex signs/sign clusters**

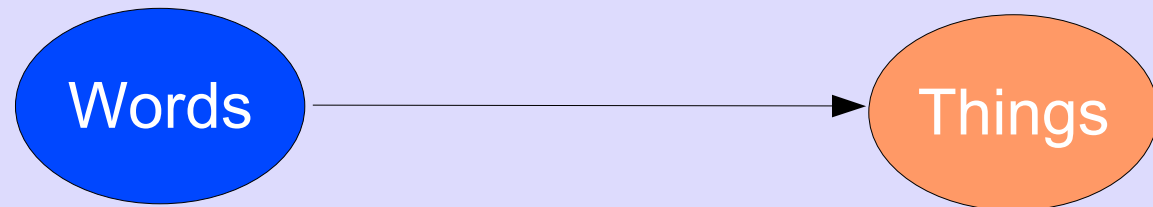
E-Science vs. E-Scholarship: different relations of research and 'documents'

- Signification and document modelling in OA related discussion up to now have basically been coined on the information model prevailing in the empirical sciences:
 - **Research** => '**Results**' => 'Packaging' => Publication
 - Robust and not very complex 'container' modell
 - Electronic Science \approx Electrified Science (e-science)
- Information modelling in the humanities and social sciences takes place in a substantially different information model:
 - **(Research \Leftrightarrow discursive 'packaging')** => Publication
 - Resulting in complex document models heavily intertwined with core research operations
 - Complex signifier \leftrightarrow significate relations as constituents.
 - 'container' models are over-reductionist and inappropriate

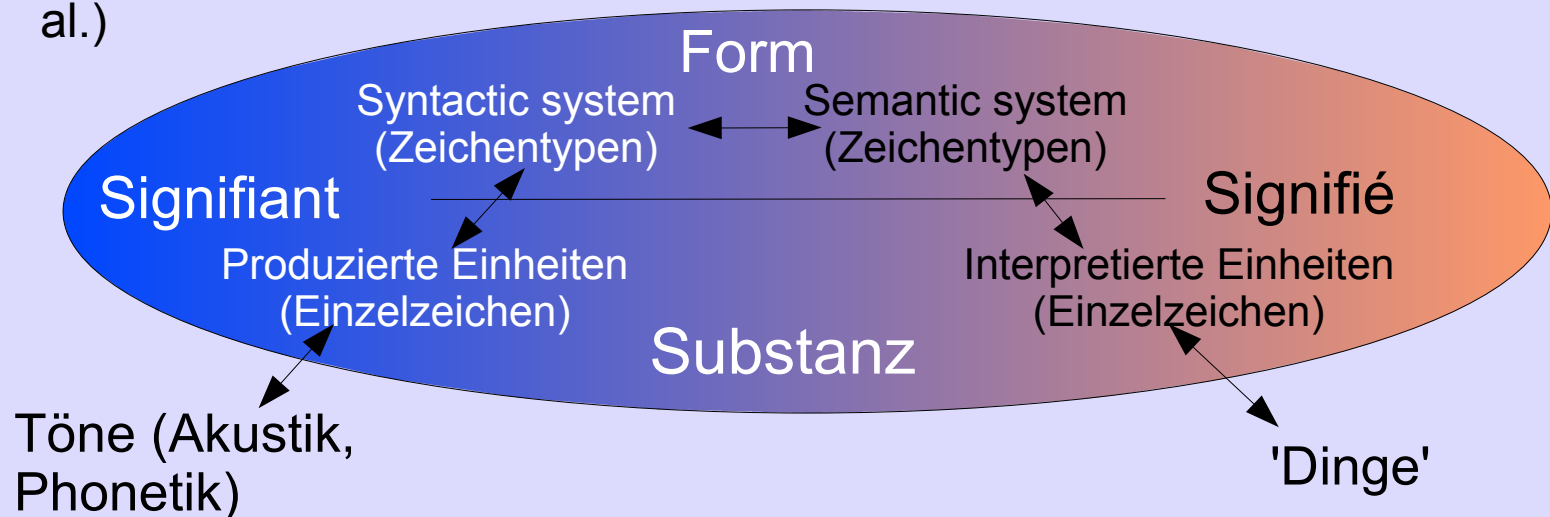
“The Pandora Box of Semiotics”

Differing modes of signification

- Discourse dominant OA and electronic publishing communities (mostly emanating from computer science) uses terms such as 'document', 'sign' or 'name' quite naively and without referring to their inherent semiological complexity.
- This results in a (technically) high level nominalist regression: the 'Pointer -> Object'-Model



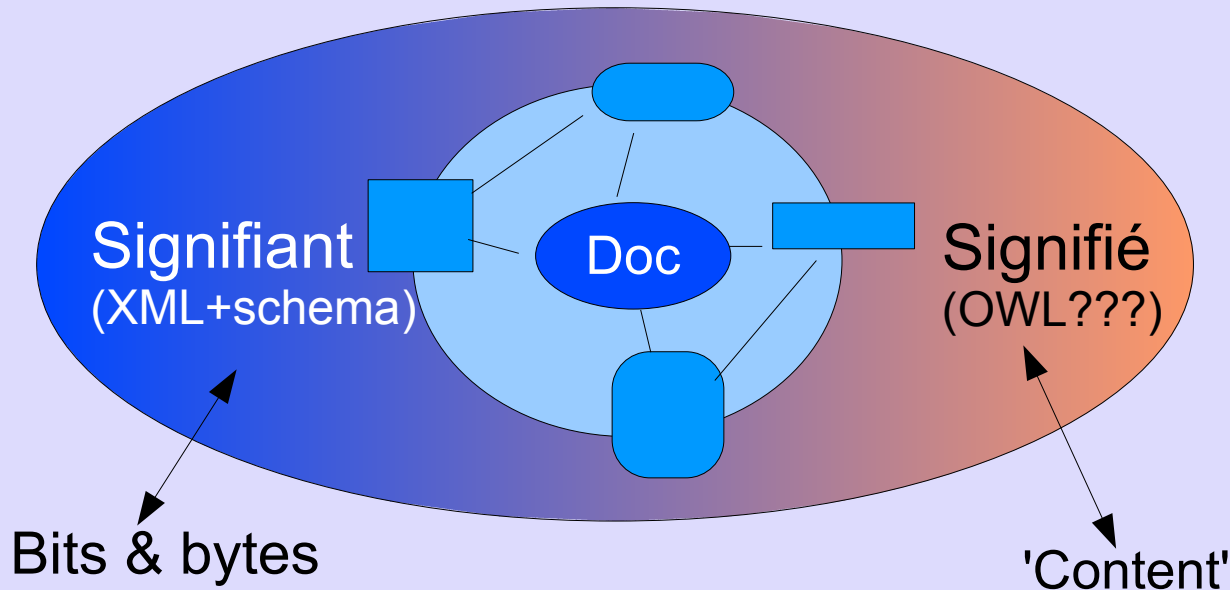
- As opposed to “the pandora box of semiotics” (de Saussure, Hjelmslev, Eco et al.)



“The Pandora Box of Semiotics”

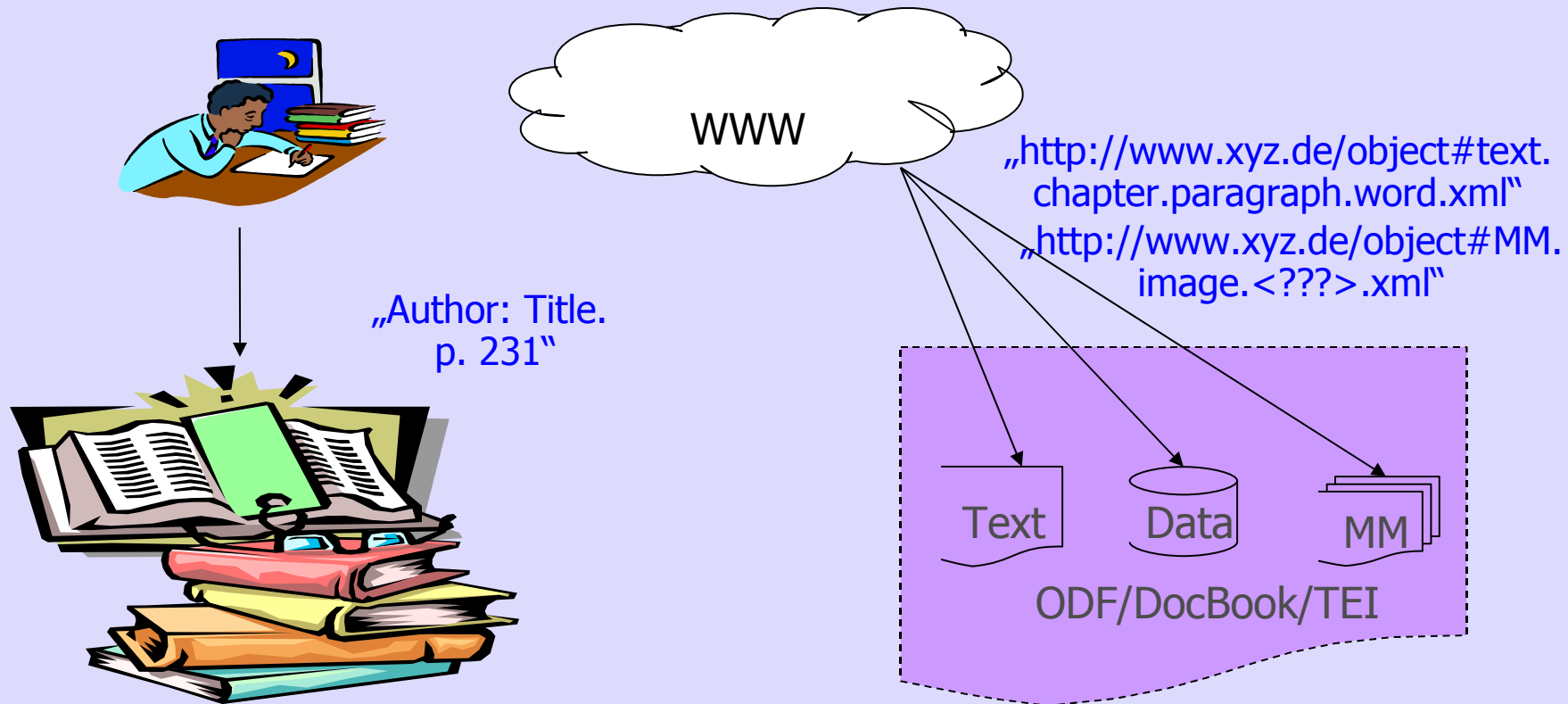
Toward a semiological de-construction of 'document'

- (electronic) text is **not just** a transcription of speech acts (parole)!
- 'text' basically remains a blank spot in linguistics and still subject to fundamental research as a complex, semiological digital object.



- => We need a model to describe documents as complex significant entities within a given pragmatic-intentional context (scholarship) ...
- => ... and thus first of all a fundamental consensus regarding the very notion of a 'document' in digital, networked settings
- => RTP-Doc (form, sign, medium) ...

Model case: quotations / references

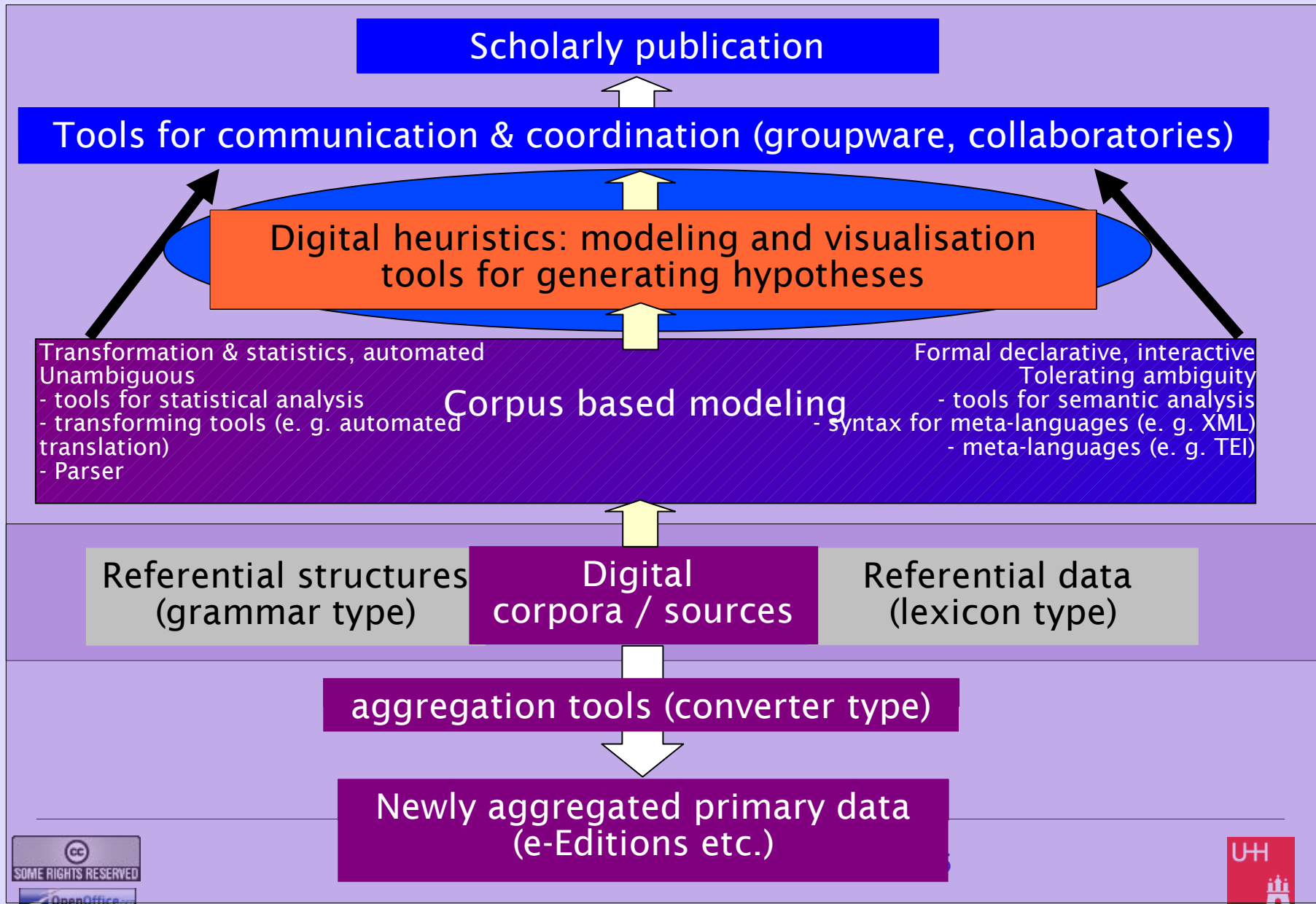


- Is identification of networked electronic documents using constructs such as DOI/URN sufficient? How to 'point' to microstructures?
- Do object models such as MPEG or TEI provide adequate conceptual frameworks?
- Replication (quotation) vs. Reference (pointer)

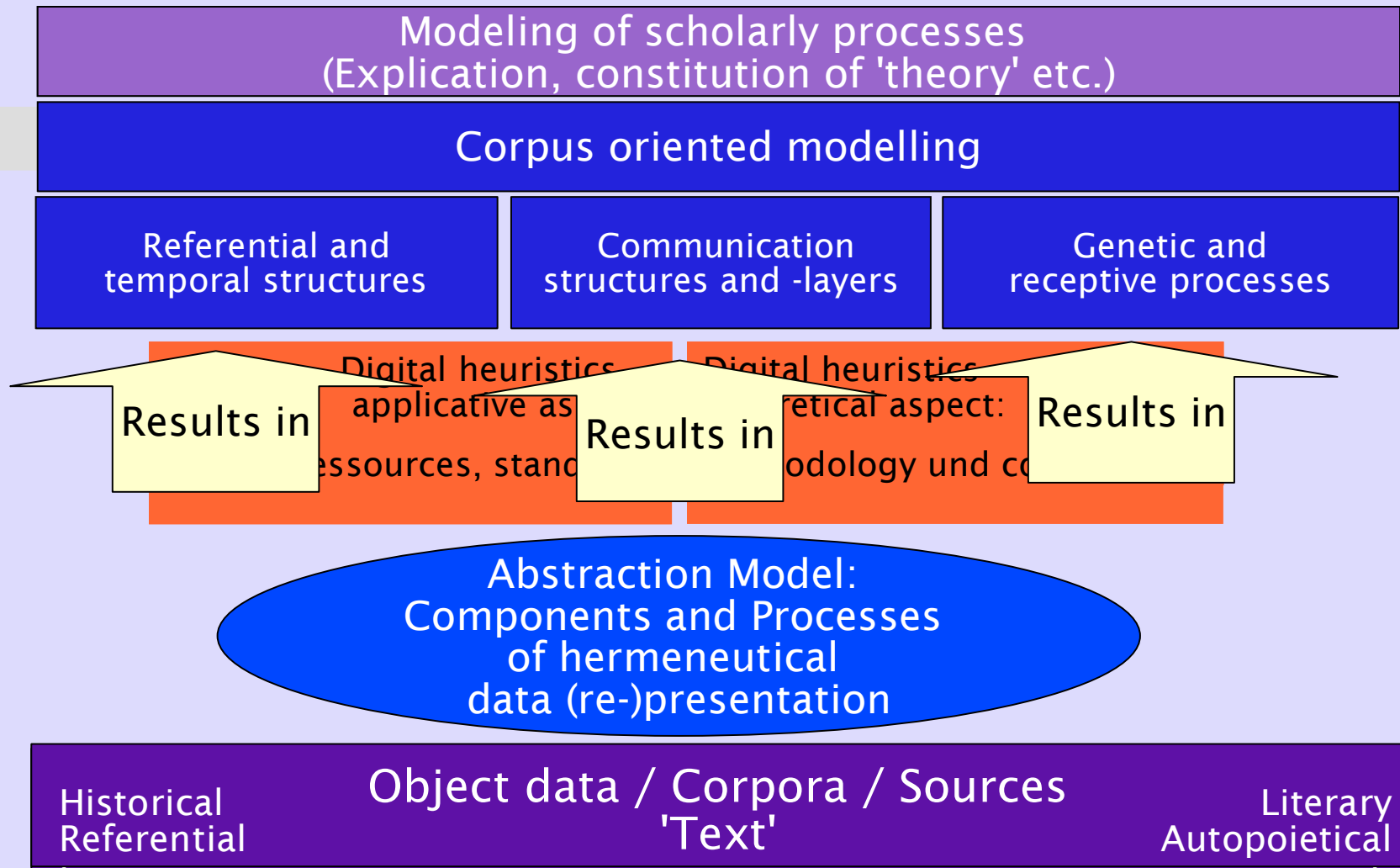
Open and standardised document models

- Digital paradigm shift in the humanities vitally depends on open and non-proprietary techniques for document modelling and authoring
- This is even more evident if one considers not just isolated documents, but webs of interrelated documents pointing and referring to each other.
- This evidence is particularly striking if one considers the need to maintain coherent webs of documents over time for decades or even centuries
- Introducing document protection technology such as for DRM in such settings would create ridiculous and nightmarish functional scenarios!

Processing of source data in the Humanities: modeling and aggregation



Constituent Layers of 'Hermeneutical Modeling' and the place of 'digital heuristics'



An example of (proto)-DH and 'Hermeneutical Modeling' environment

EU-NSF-Projekt Cultural Heritage Languages Technology (CHLT)*

*Thanks to Bruce Fraser (Cambridge) and other participants in CHLT!

- Greek Lexicon Project (Cambridge): TGL => LSJ => TLG => GLP
- XML online-slips + XSLT
 - Supports annotation
 - Can be organised in multiple dimensions
 - Building block of a distributed, networked 'Collaboratory' (together with Perseus DL and others) for cultural heritage research work based on advanced technology
- Combined with tools for lexical clustering and visualisation of lexical distribution this is evolving into a fertile environment for generating text-related hypotheses.
- Illustrates well the methodological advantages of a clearly delimited and highly 'pre-aggregated' corpus.

Lexicon Code Fragment

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet href="../../Lexicon/Dtd/lexicon.css" type="text/css"?>
<!DOCTYPE lexicon SYSTEM "file:///../../Lexicon/Dtd/lexicon.dtd">
<lexicon>
  <header>
    <file> <title>Greek Lexicon Sample Page 1</title><editor>AAT</editor><date>7/4/04</date></file>
  </header>
  <text>
    <body>
      <ANE><HG><HL>λαγω<hyph/>βόλον</HL><VL><Lbl>also</Lbl><FmHL>λαγωβόλον</FmHL><Au>Anth.</Au>
        </VL><Infl>ου</Infl><PS>n</PS><Ety><Ref>λαγώς</Ref>,
          <Ref>βόλος</Ref></Ety></HG>
          <S1><Qualif>orig.</Qualif><Def>stick for throwing at hares<Expl>in
            hunting; or simply as a mark of the
            countryman</Expl></Def><S2><Tr>throwing-stick, stick</Tr><Au>Theoc.
              Anth.</Au></S2>
          </S1></ANE>
      <ANE><HG><HL>λαγῶδιον</HL><Infl>ου</Infl><PS>n</PS></HG>
        <S1><Def>young hare</Def><Tr>leveret</Tr><Au>Ar.</Au>
        </S1>
      </ANE>
      <ANE><HG><HL>λαγών</HL><Infl>όνος</Infl><PS>f</PS><Ety>reltd.
        <Ref>λαγαρός</Ref></Ety></HG>
        <S1><Tr> flank, side, waist<Expl>of a person or animal, ref. to the
          area betw. the ribs and the hip, or more generally, in sg. or pl., to the
          middle of the body</Expl></Tr><Au>E. <NBPlus/></Au><S2><Tr>side <Expl>of a
            mountain, a river</Expl></Tr><Au>Call. Anth.</Au></S2>
        </S1>
        <S1><Tr>recess, hollow<Expl>of a container, such as a cup, a
          quiver</Expl></Tr><Au>Eub. Anth.</Au><S2><Indic>under an overhanging
            rock</Indic><Au>Plu.</Au></S2>
        </S1><Ann><Para>how to translate Call.5.88 breasts and hips of Athena?
          or lagones more generally for the body, or the middle area of the body; perh.
          waist here. Sense 'womb', see Rev.Suppl., prob. doesn't exist. </Para><Para>at
            [...]
```

body VE vHG HL λιβάζομαι HL PS mid.vb.

PS Etymology [Ref λίβας Ref] Etymology vHG

vS1 Indic (of a fountain) Indic Def gush with water

Def Tr flow Tr Au Anthol. Au vS1 VE

ANE HG HL λιβανίδιον HL Infl ου Infl PS

f. PS Etymology [dimin. Ref λίβανος Ref] Etymology HG

S1 Tr little bit of incense Tr Au Men. Au

S1 ANE

ANE HG HL λίβανος HL Infl ου Infl PS f.

rom (also Ital m. Ital sts. in sense 2) rom.

PS Etymology [Semit. loanwd.] Etymology HG

S1 Nm 1 Nm Tr frankincense-tree Tr Au

Hdt. Thphr. Au S1

S1 Nm 2 Nm Def aromatic gum resin Expl (fr.

the frankincense-tree, either the raw product or the smoke when it is burned) Expl

λαγω-βόλον, also **λαγωβοβόλον** Anth. ου *n.* [λαγώς, βόλος] orig., stick for throwing at hares (in hunting; or simply as a mark of the countryman); **throwing-stick, stick** Theoc. Anth.

λαγώδιον ου *n.* young hare, leveret Ar.

λαγών όνος *f.* [reld. λαγαρός] **1 flank, side, waist** (of a person or animal, ref. to the area betw. the ribs and the hip, or more generally, in sg. or pl., to the middle of the body) E. +; **side** (of a mountain, a river) Call. Anth.

2 recess, hollow (of a container, such as a cup, a quiver) Eub. Anth.; (under an overhanging rock) Plu.

λαγῶς ου *adj.* [λαγώς] **of a hare** (ref. to the meat) Ar. || NEUT.PL.SB. (w. κρέα understood) hare-meat, cooked hare-meat dish Ar.

λαγῶς λαγῶ, also **λαγῶς** λαγῶ, ep. **λαγῶς** οὔ Ion. **λαγός** οὔ *m. and f.* | acc.sg. λαγών (Ar.) |

1 hare Hom. +; (as a type of timidity or cowardice, esp. in provbl.phrs.) Posidipp. D. +

λαθι-κηδής, Aeol. **λαθικᾶδής** ές *adj.* [λανθάνω, κήδος] | acc.sg. λαθικᾶδεον | causing forgetfulness of care; **soothing** —of a mother's breast ll.; **banishing cares or pain** —of wine Alc. —of Apollo (as healer), of medical knowledge Anth.

λαθί-πονος ου *adj.* **forgetful of pain or trouble** S.

λαθι-πορφυρίς ίδος *f.* a kind of bird (app. which is hard to see, or is active only at night), perh., **shy purple-gallinule** lbyc. | see also πορφυρίς

λαθί-φθογγος ου *adj.* causing forgetfulness of speech, **silencing voices** —of death Hes.Sc.

λαθιφροσύνη ης *f.* [reld. φρονέω] (pl.) forgetfulness of common-sense, **madness** A.R.

λαθοίατο (aor.2 mid.optat.): see λανθάνω

λᾶθος εος *dial.n.* [reld. λήθη] **forgetfulness, indifference** (as the cure for love) Theoc.

λαγών ὄνος *f.* [reltd. λαγρός]

- how to translate Call.5.88 breasts and hips of Athena? or lagones more generally for the body, or the middle area of the body; perh. waist here. Sense 'womb', see Rev.Suppl., prob. doesn't exist.
- at Theoc.22.202 the spear pierces the lagwn and the omphalos: side/midriff and navel? Gow: the unprotected part of the abdomen between ribs and hips
- Plu.Arat.22: is lagwn here a hollow, recess, or just the side of a mountain? Are there really two senses for mountains? sense 2 seems to be needed for the cup and quiver.

λαγῶς ον *adj.* [λαγός]

- does uncontr. lagwios exist neut.pl.sb., Ar.V.709 -- delicacies as LSJ, or every kind of hare dish, plenty of hare meat

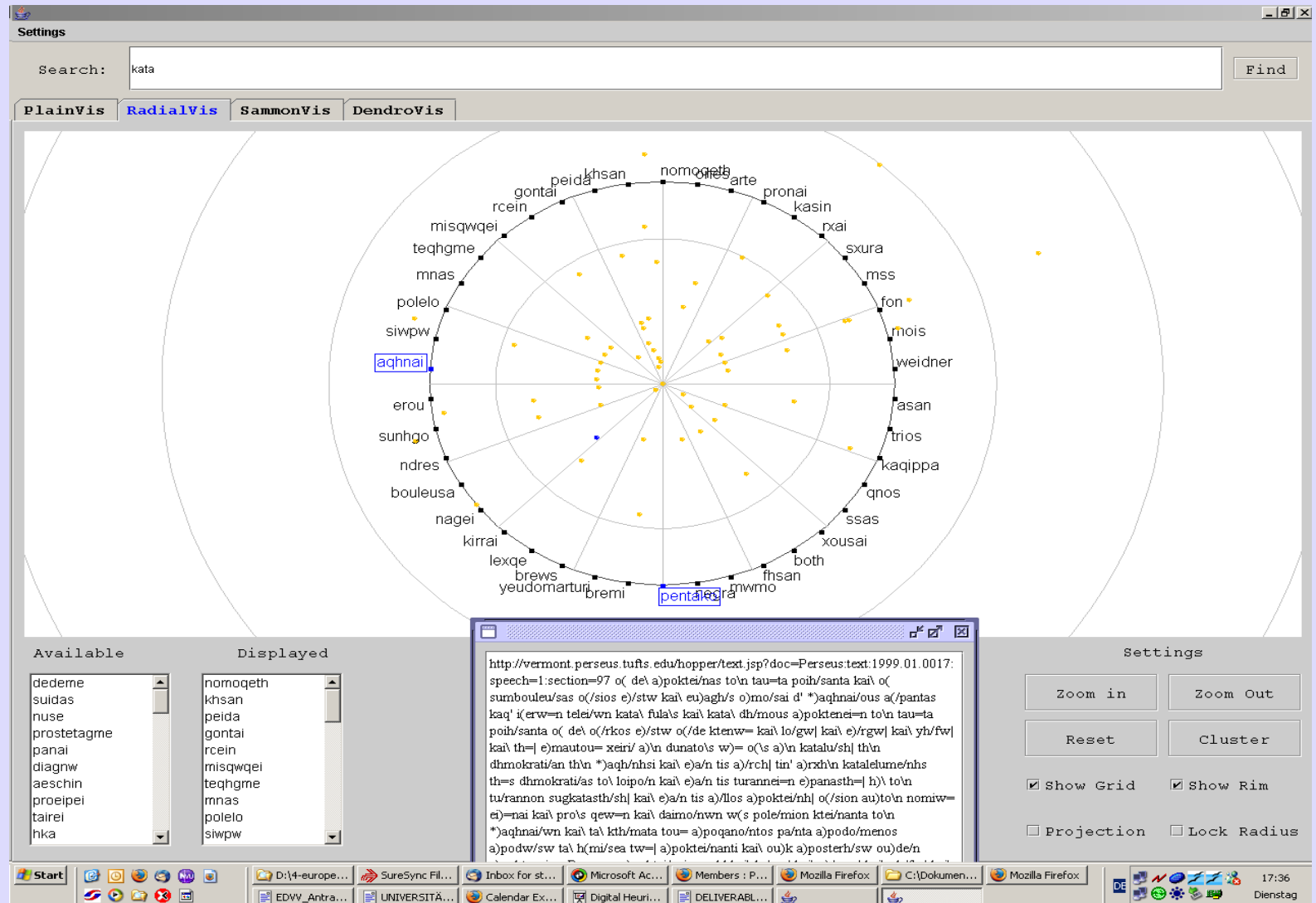
λαγῶς λαγῶ., also **λαγῶς** λαγῶ, ep. **λαγῶς** οὔ, Ion. **λαγός** οὔ *m. and f.* | acc.sg. λαγών (Ar.) |

- does this include rabbit? kuniklos is late. (mod. kouneli; lakoudaki is bunny i.e.rabbit, not small hare?)

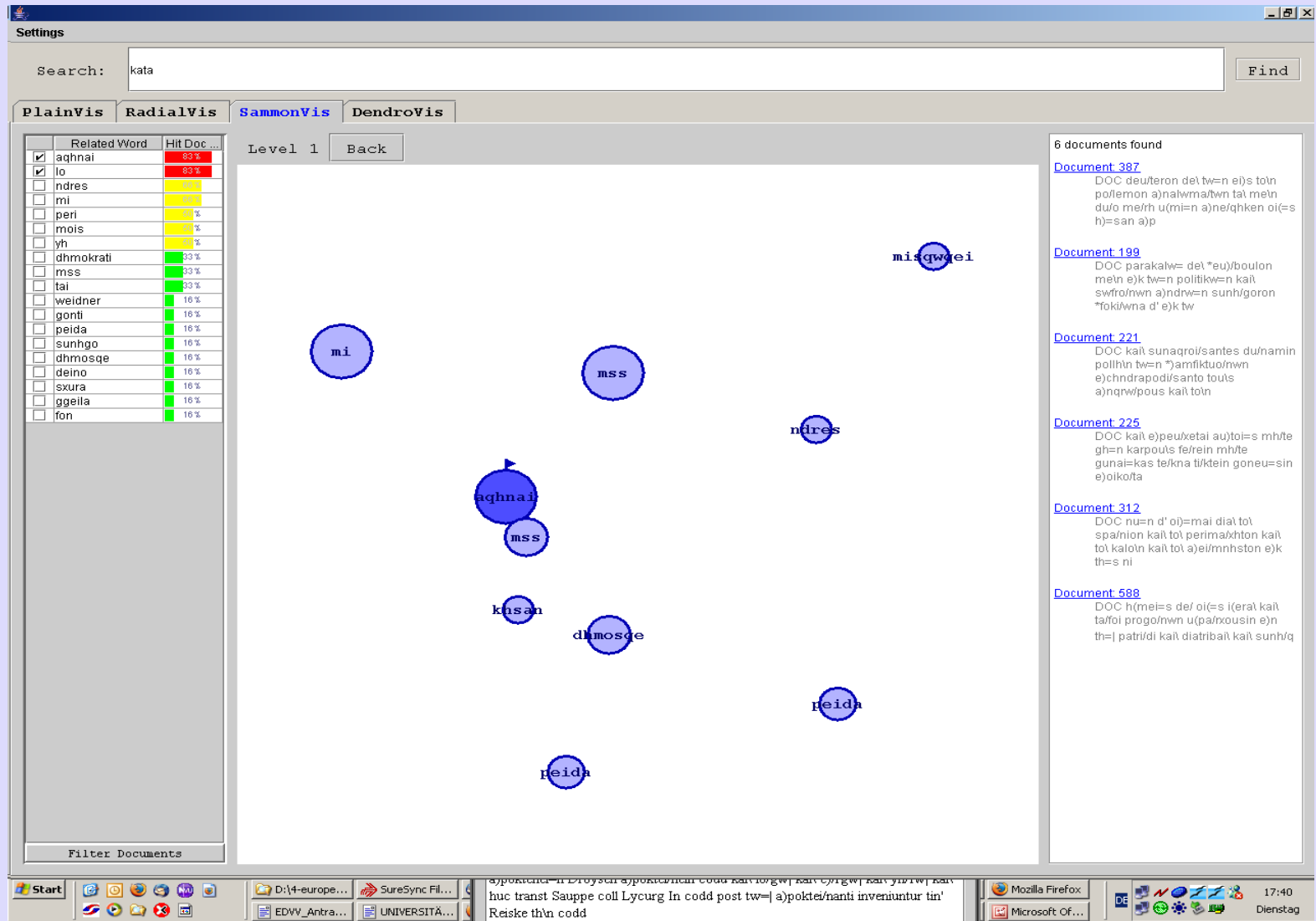
(quickly?) and put it on. i.e. intensive of a diff. sense of lamabanw from grip a tool etc.

- it's difficult to assess how many exx. have the sense of 'pick up, take' and how many are just 'hold'.
- perh. sense 2 is wrong, it just means 'take, accept, pick up' and there is no intensive (eagerly or sim.). The garment ex. (Theoc.15.21) could then go here, pick up the garment (in order to put it on to go out) 1 and 2 could perhaps all be combined as one section
- Ar.Lys.209 sexual double-entendre here? most of the uses of this vb. would suit this. But in which section shd. this passage be placed? w.gen., construction needs to be pointed out -- but sense: each woman holds on to one part of the kulix, or each holds it in turn, or each drinks some of the wine in turn??
- There seems to be a connection betw. this ctxt. and Theoc.18.46: we will first draw fr. the silver flask and let drip smooth oil beneath that shady plane. (Gow.) Do they take the ointment \ or oil from the olpis into their hands, and then smear / or pour it in drops ? Or does the taking just loosely ref. to picking up the olpis along with the aleiphar.
- Does this passage lead us to think the Ar. passage

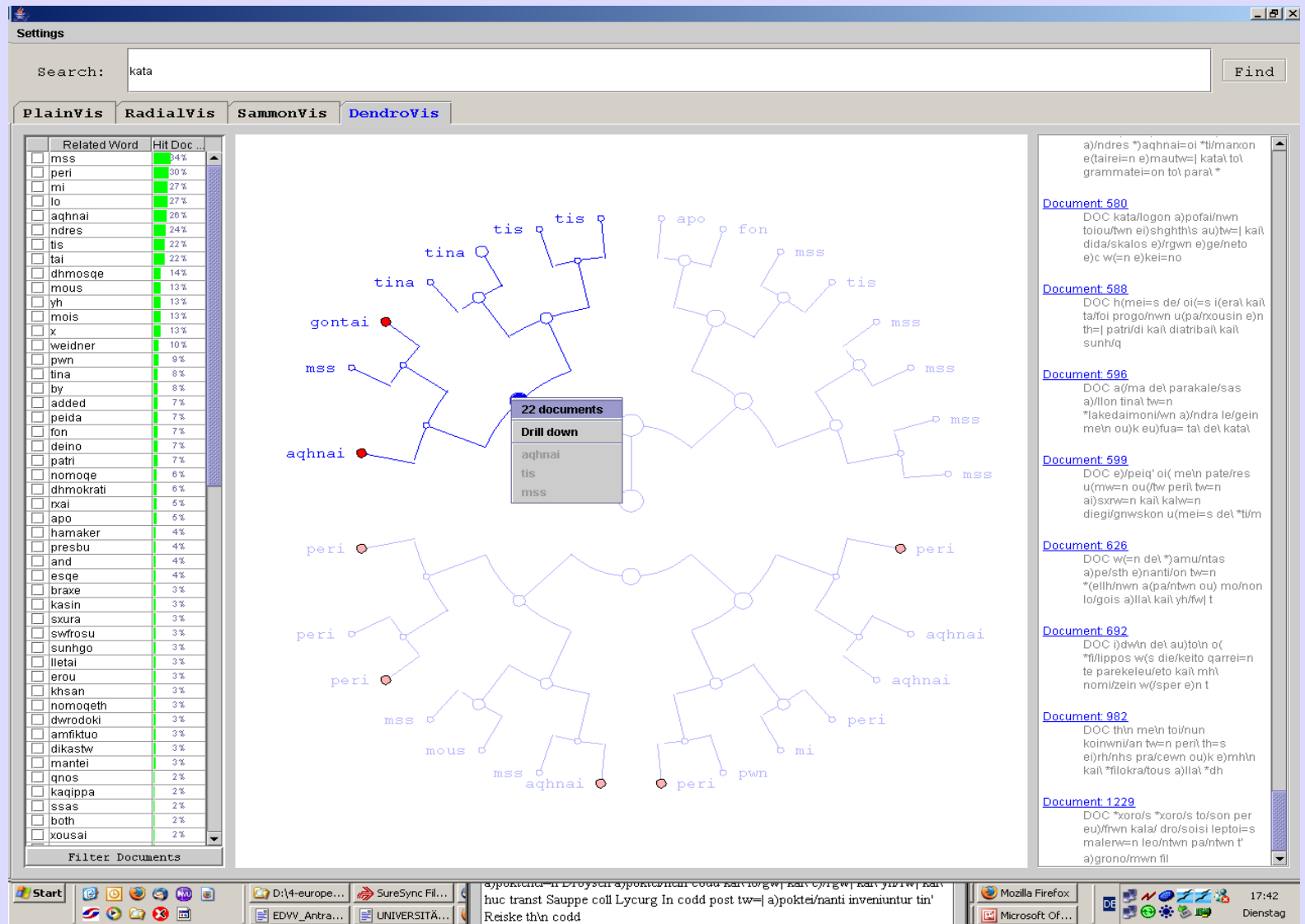
Lexicon <=> Corpus Visualisation I



Lexicon <=> Corpus Visualisation II



Lexicon <=> Corpus Visualisation III



Conclusions I: the specific requirements of eScholarship w. r. t. Open Source and Open Access

- For innovative processing of digital sources to work at all a very specific understanding of the term 'open source' needs to be consequently and systematically applied: free availability of all source material!
 - Hence the primary characteristic of cyberinfrastructure as seen by the ACLS: *"It will be accessible as a public good"*
- The heuristics used for novel corpus modelling as well as their technical implementations and foundations need to be open source in the more traditional sense of the term as well as based on open standards to enable future digital hermeneutical heuristics
 - Hence ACLS recommendation 7: *"Develop and maintain open standards and robust tools"*

Conclusions II: eScholarship in the broader context of Open Source and Open Access communities

- The mainstream of the e-science OA discussion almost completely bypasses e-scholarship in that questions of publication economy and access to journal articles are of minor relevance in our sector
- Instead, the need for OS and OA in e-scholarship stems from the needs of the rapidly evolving working paradigm of digital scholarship: OS and OA are key enablers for this new paradigm!
 - Hence recommendation 2 of the ACLS report: *“Develop public and institutional policies that foster openness and access.”*
- OA discussions in e-scholarship and e-science share concerns with key issues such as quality assurance and reputation building – but going into these would definitely be beyond the time scope of this paper.
 - Thank you for your patience and attention